Time Series Regression Analysis to Forecast Bike Rental Demands and Analysis of Bike Rental Patterns Using JMP® 11

Zabiulla Mohammed, Graduate Student, Oklahoma State University

Vandana Reddy, Graduate Student, Oklahoma State University

Abstract

In today's world, with environmental and health issues gaining significant importance, commuting by ecofriendly modes of transportation is becoming popular. One such environment friendly means which is gaining popularity in the USA is biking activity. With the increase in the number of bike enthusiasts, more and more bike renting companies are trying to find their foothold in this market. One challenge these companies face is to predict the number of casual customers, who rent a bike occasionally and the number of registered customers, who rent more regularly and have a membership with the company. It is also important for these rental companies to understand the renting patterns of different customers over time.

Objective

Our objective is to understand the bike rental patterns of the casual and registered users based on two years of rental data and use JMP Pro 11's Time series analysis feature to forecast the number of casual and the number of registered users for 25 weeks.

Data Preparation

The bike rentals dataset gathered from UCI Machine Learning database has various input variables related to weather conditions, holidays, temperature, working day, season, humidity and so on. The two target variables chosen are *'the number of casual users'* and *'the number of registered users'* of bikes.

Some patterns

1. Casual rental users tend to rent a bike more often in the afternoons (Fig1.1)
2. Registered users rent a bike more often during the beginning of office hours (7 am – 9 am) and after the office hours (5 pm – 8pm) (Fig1.2)

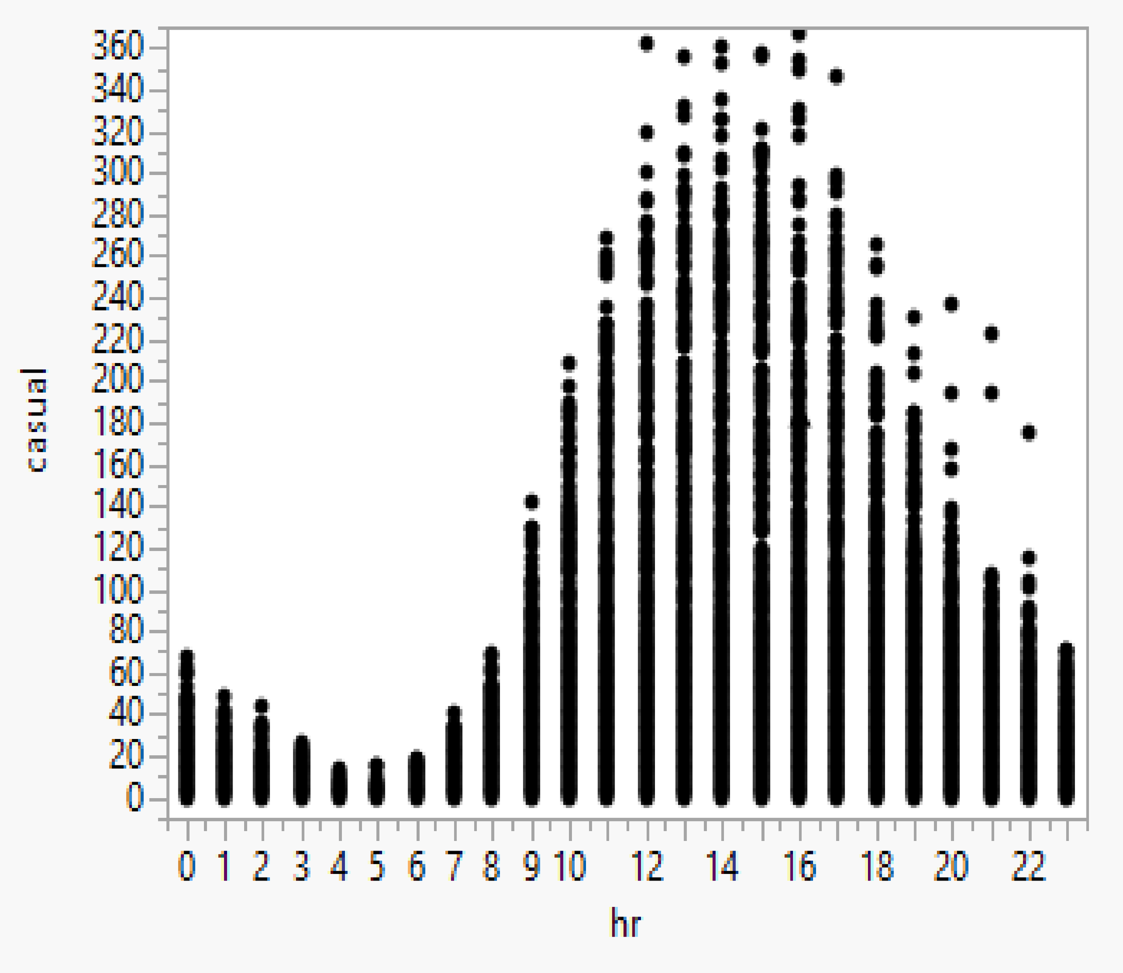


Fig 1.1

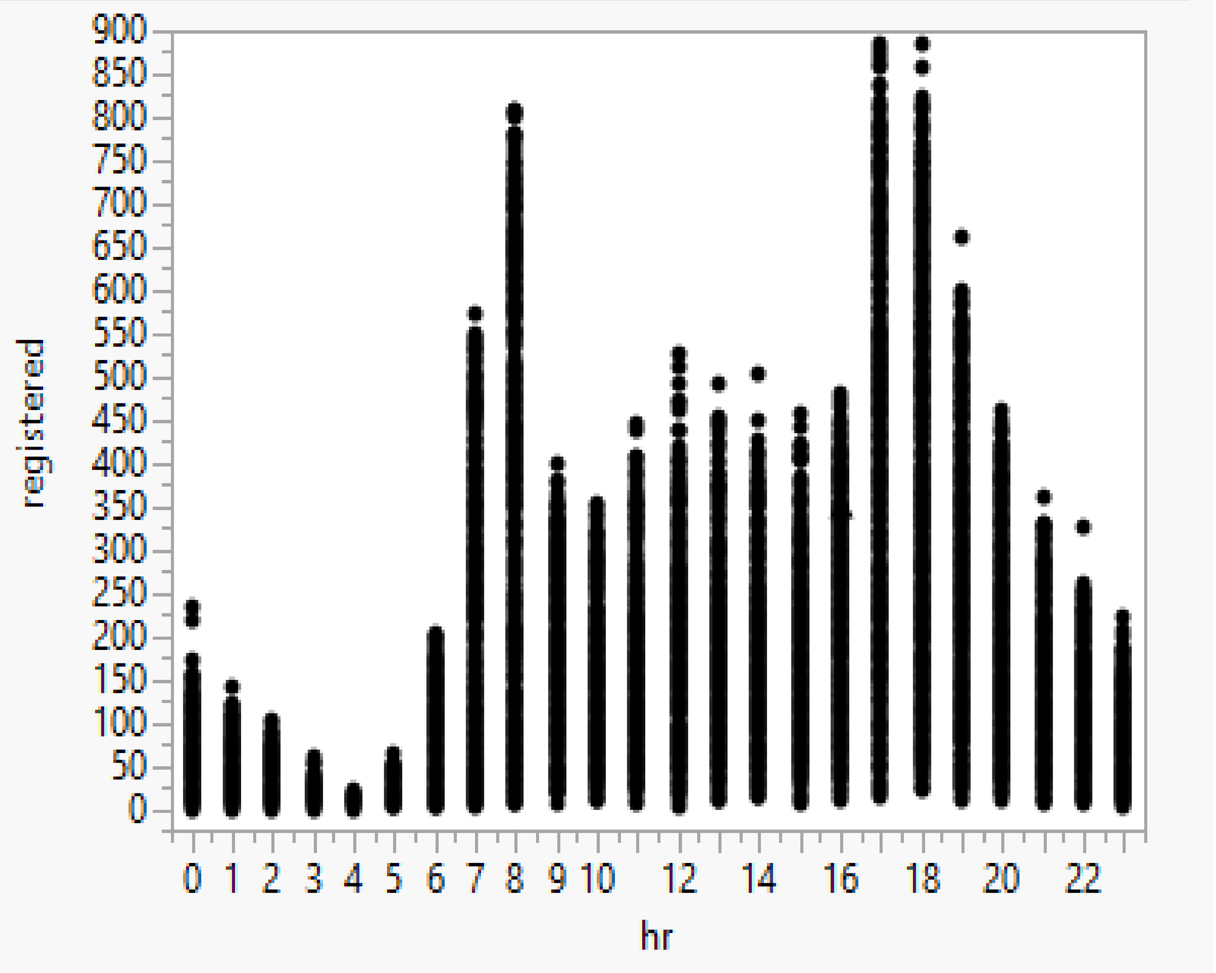


Fig 1.2

3. Casual rentals are more in number on the weekends and non-working days than on working days. (Fig2.1)
4. Registered rentals are more often during the weekdays and working days (Fig2.2)

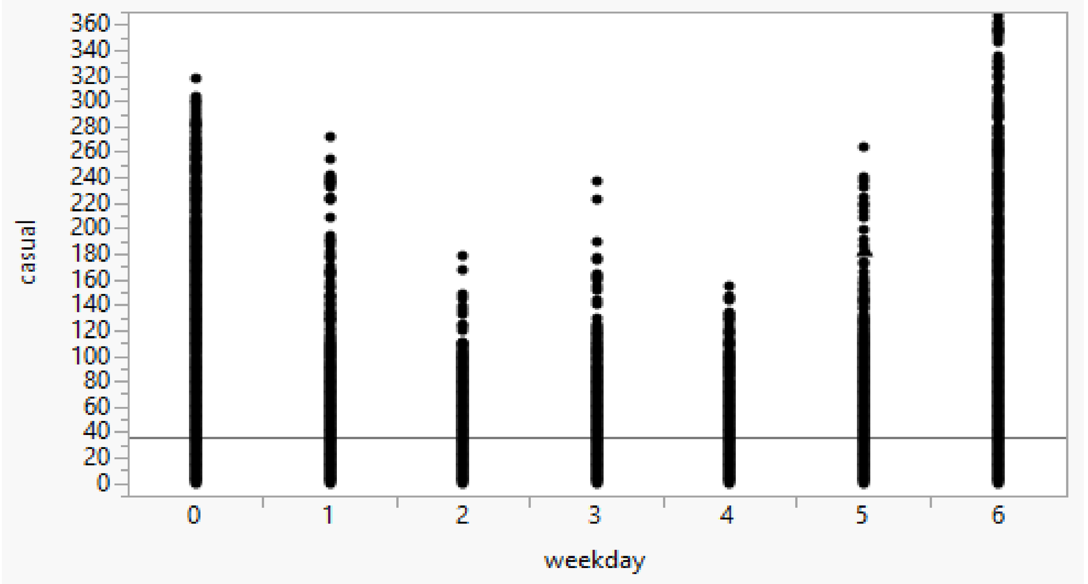


Fig 2.1

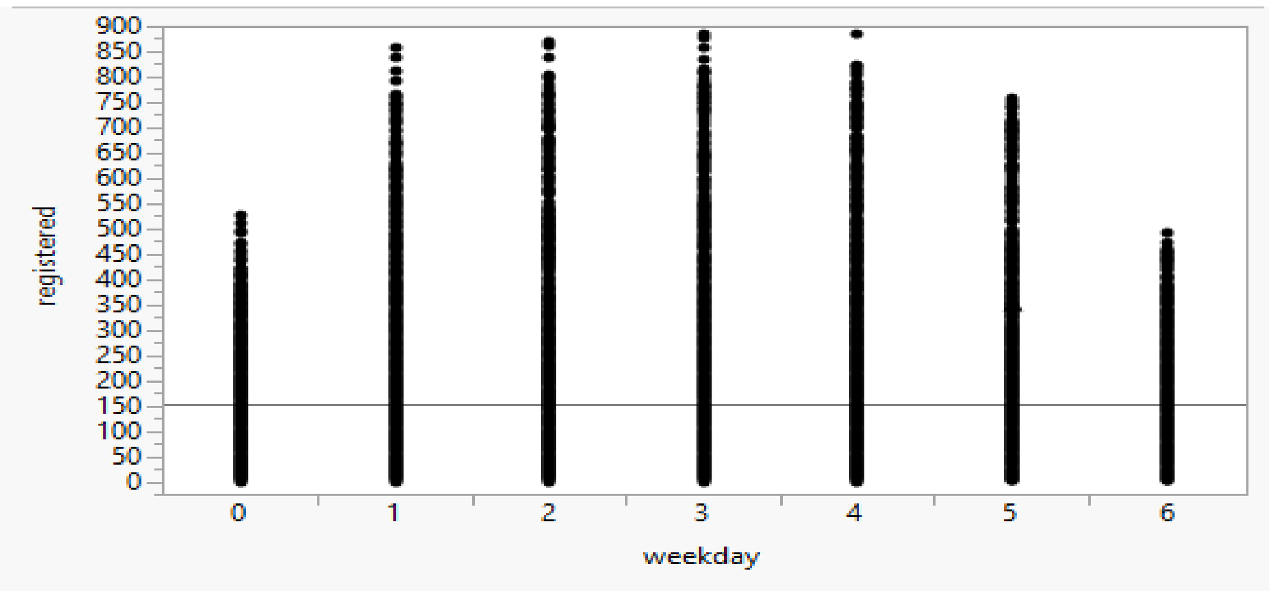


Fig 2.2

Linear Regression

Two linear regression models were built to predict the number of casual and registered renters using holiday, working day, average temperature, humidity and wind speed as predictor variables.

Target: Casual Renters

Summary of Fit

RSquare	0.674346
RSquare Adj	0.671647
Root Mean Square Error	393.449
Mean of Response	848.1765
Observations (or Sum Wgts)	731

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	792.36128	99.50154	7.96	<.0001*
yr	280.58598	29.35746	9.56	<.0001*
holiday	-301.72	90.07202	-3.35	0.0009*
workingday	-832.3999	32.39132	-25.70	<.0001*
atemp	2368.6213	91.62695	25.85	<.0001*
hum	-711.3427	106.8933	-6.65	<.0001*
windspeed	-963.6506	196.471	-4.90	<.0001*

Target: Registered Renters

Summary of Fit

RSquare	0.731128
RSquare Adj	0.7289
Root Mean Square Error	812.383
Mean of Response	3656.172
Observations (or Sum Wgts)	731

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	1484.5594	205.4481	7.23	<.0001*
yr	1729.1464	60.61649	28.53	<.0001*
holiday	-275.0226	185.9783	-1.48	0.1396
workingday	913.40499	66.88074	13.66	<.0001*
atemp	4748.6587	189.1889	25.10	<.0001*
hum	-1572.842	220.7105	-7.13	<.0001*
windspeed	-3023.7	405.6681	-7.45	<.0001*

The following have been observed using the two models above:

1. Holidays are significant for casual renters whereas for registered users holidays do not matter.
2. Working day has a negative estimate for casual renters and positive estimate for registered renters which means that on a working day, the number of casual rentals go down by 832 units whereas the number of registered rentals go up by 932 units keeping all the other factors constant.
3. Higher humidity and wind speed levels generally tend to reduce the number of rentals for both casual and registered users.

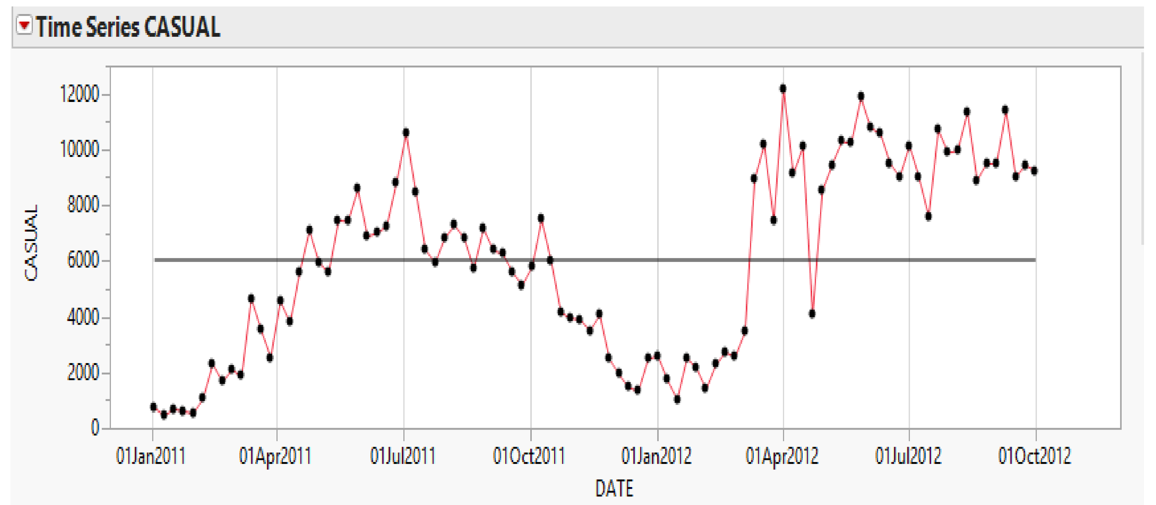
Time Series Analysis for forecasting bike users

The data over two years is aggregated weekly to form 104 rows each row representing a week. Each row has a date or timestamp, number of casual bike renters on that day and the number of registered users who rented a bike on that day. This data has been divided such that the last 12 weeks of data is used as holdout to assess model performance. The rest 92 weeks of data is used in building models.

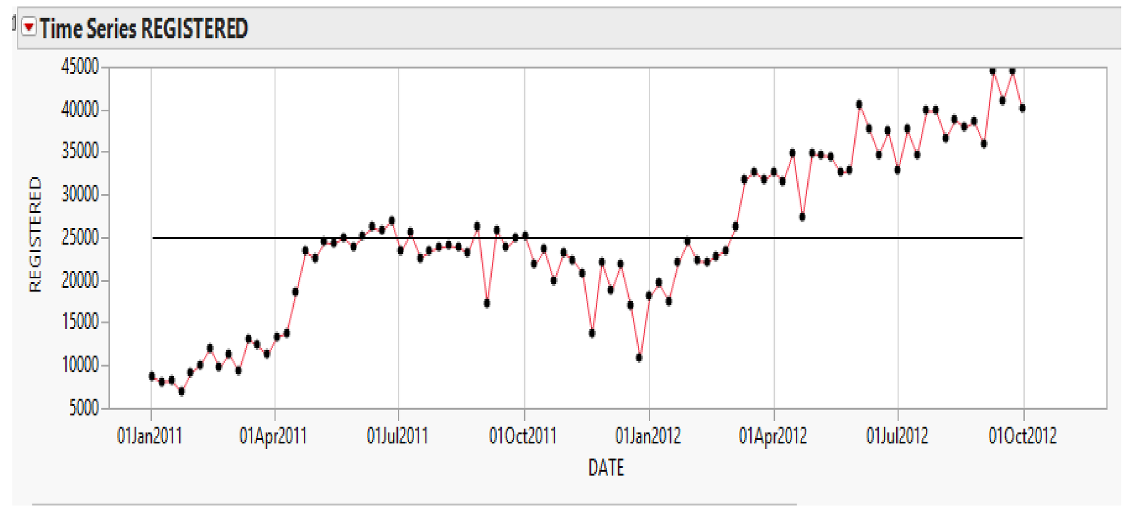
Time Series Regression Analysis to Forecast Bike Rental Demands and Analysis of Bike Rental Patterns Using JMP® 11

Zabiulla Mohammed, Graduate Student, Oklahoma State University
Vandana Reddy, Graduate Student, Oklahoma State University

Time series for casual users



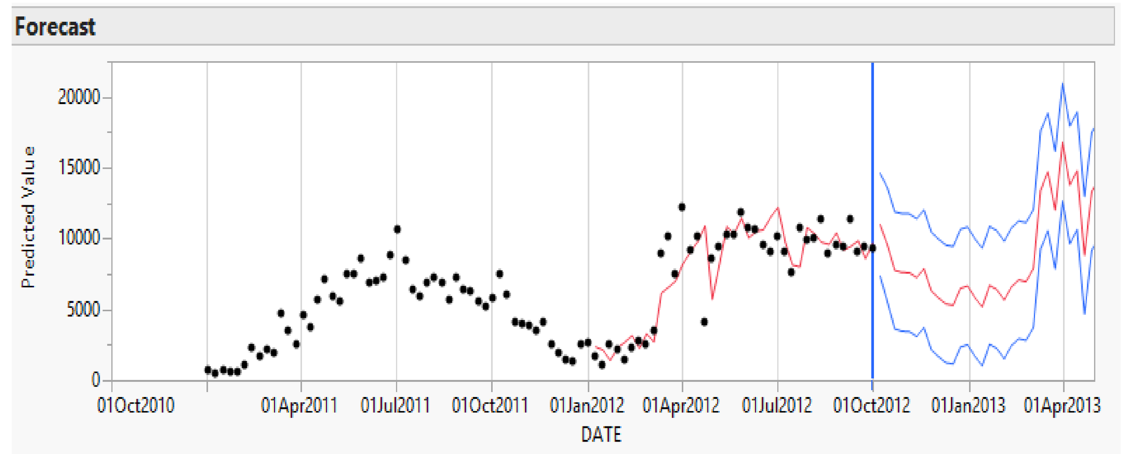
Time series for registered users



ARIMA Model group comparison and forecast graph for casual users

Model Comparison

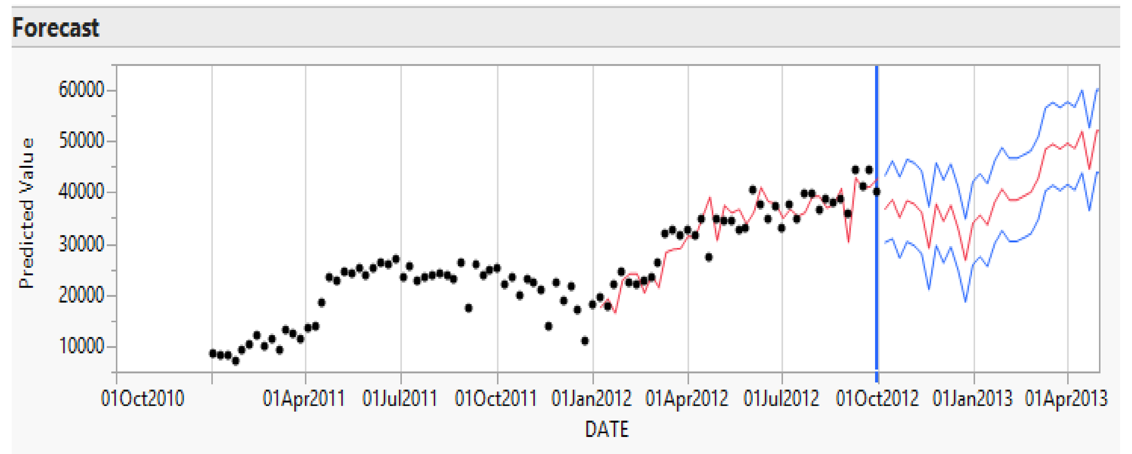
Report	Graph	Model	DF	Variance	AIC	SBC	RSquare	-2LogLH	Weights	AIC Rank	MAPE
✓	✓	Seasonal ARIMA(1, 1, 1)(0, 1, 0)52	36	3406281.5	702.82033	707.81102	0.710	696.82033	0.418954	1	23.551044
		Seasonal ARIMA(1, 1, 1)(1, 1, 0)52	35	904.18614	704.82032	711.47457	0.710	696.82032	0.154126	2	23.551035
		Seasonal ARIMA(1, 1, 1)(0, 1, 1)52	35	1751803.6	704.82033	711.47457	0.710	696.82033	0.154125	3	23.551040
		Seasonal ARIMA(1, 1, 1)(1, 1, 1)52	34	0.0145829	706.82032	715.13813	0.710	696.82032	0.056700	4	23.550950
		Seasonal ARIMA(0, 1, 1)(0, 1, 0)52	37	4140260	706.97269	710.29981	0.658	702.97269	0.052540	5	25.411277
		Seasonal ARIMA(1, 1, 0)(0, 1, 0)52	37	4241756.1	707.84310	711.17022	0.650	703.8431	0.034001	6	26.530670
		Seasonal ARIMA(0, 1, 0)(0, 1, 0)52	38	4397096.2	708.22590	709.88947	0.628	706.2259	0.028078	7	26.899080
		Seasonal ARIMA(0, 1, 1)(1, 1, 0)52	36	4255267.3	708.97269	713.96337	0.658	702.97269	0.019329	8	25.411277
		Seasonal ARIMA(0, 1, 1)(0, 1, 1)52	36	4255267.3	708.97269	713.96337	0.658	702.97269	0.019329	9	25.411277



ARIMA Model group comparison and forecast graph for registered users

Model Comparison

Report	Graph	Model	DF	Variance	AIC	SBC	RSquare	-2LogLH	Weights	AIC Rank	MAPE
✓	✓	Seasonal ARIMA(1, 1, 1)(0, 1, 0)52	36	11154592	748.83082	753.82150	0.768	742.83082	0.186971	1	8.296188
		Seasonal ARIMA(0, 1, 1)(0, 1, 0)52	37	12112628	748.86918	752.19631	0.756	744.86918	0.183418	2	8.095881
		Seasonal ARIMA(1, 1, 0)(0, 1, 0)52	37	12384958	749.68105	753.00817	0.750	745.68105	0.122222	3	8.277180
		Seasonal ARIMA(1, 1, 1)(0, 1, 1)52	35	5736654.1	750.83071	757.48496	0.768	742.83071	0.068786	4	8.296141
		Seasonal ARIMA(1, 1, 1)(1, 1, 0)52	35	32377.329	750.83103	757.48528	0.768	742.83103	0.068775	5	8.296281
		Seasonal ARIMA(0, 1, 1)(0, 1, 1)52	36	12449090	750.86918	755.85987	0.756	744.86918	0.067476	6	8.095881
		Seasonal ARIMA(0, 1, 1)(1, 1, 0)52	36	12449090	750.86918	755.85987	0.756	744.86918	0.067476	6	8.095881
		Seasonal ARIMA(1, 1, 0)(1, 1, 0)52	36	12728985	751.68105	756.67173	0.750	745.68105	0.044963	8	8.277180
		Seasonal ARIMA(1, 1, 0)(0, 1, 1)52	36	12728985	751.68105	756.67173	0.750	745.68105	0.044963	9	8.277180
		Seasonal ARIMA(0, 1, 0)(0, 1, 0)52	38	13451875	751.83470	753.49826	0.722	749.8347	0.041638	10	8.677494
		Seasonal ARIMA(1, 1, 1)(1, 1, 1)52	34	4982.002	752.83060	761.14841	0.768	742.8306	0.025306	11	8.296094
		Seasonal ARIMA(0, 1, 1)(1, 1, 1)52	35	12804778	752.86918	759.52343	0.756	744.86918	0.024823	12	8.095881



Model Selection

- The best models selected for casual and registered users using the subgrouping feature of JMP are the ARIMA (1,1,1) Seasonal ARIMA (0,1,0).
- These models chosen have the lowest AIC values. JMP simplifies model selection by calculating AIC ranks.
- The forecast values are compared by saving the time series model prediction values and comparing them with the values in the holdout sample of 12 weeks. The average MAPE on this hold out sample for casual users is 23.55 and for registered rentals is 8.296.

Recommendations

- Since registered users rent during the rush hours (7 am-9 am and 5 pm – 7 pm) and casual users rent during working hours (10 am – 7 pm) it is suggested to maintain inventory levels appropriately to meet the rental demands.
- To attract more casual users it is recommended to launch promotions during holidays and weekends.
- Bike rental companies can anticipate a decline in business during bad weather.
- The ARIMA models built can be used to forecast the bike rental demands for the casual and registered users for the next 25 weeks.

References

https://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset
Fanaee-T, Hadi, and Gama, Joao, 'Event labeling combining ensemble detectors and background knowledge', Progress in Artificial Intelligence (2013): pp. 1-15, Springer Berlin Heidelberg
http://support.sas.com/documentation/onlinedoc/jmp/11/UsingJMP.pdf

Acknowledgement

We thank Dr. Goutam Chakraborty, Professor (Marketing), Spears School of Business and founder of SAS and OSU Data Mining Certificate Program – Oklahoma State University for his encouragement and support.